



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

A COMPARISON OF A GIRLS' REFORM SCHOOL, ATTENDANTS AT A STATE HOSPITAL FOR THE INSANE, AND PUBLIC SCHOOL CHILDREN, BY MEANS OF CERTAIN TESTS OF INTELLIGENCE¹

S. L. PRESSEY²

I. NEED FOR THE MEASUREMENT OF "APPLICABILITY" IN USING SCALES OF INTELLIGENCE

The writer has, for a number of years, been much interested in the question as to how far various scales for the measurement of intelligence can be considered applicable to adults, and to atypical and pathological cases. The present paper, it should be said at once, is much more of a study of the tests concerned than of the cases worked with. In previous articles³ the writer has tried to show that, because of the influence of maturity, special features of environment, etc., which affect performance on various tests in various ways, only a rough reliability can be expected of either the Binet or the Yerkes Point Scale in work with adults and with psychotic (insane) cases. The accumulation of data, with regard to such problems, by means of scales which must be given individually is, however, a long and laborious process. The problem may be quite as well come at by means of group scales; a much larger mass of data may be thus worked up in a given amount of time—and the writer has a private conviction that, for research purposes at least, the group tests give more satisfactory results than the individual examination would give. That is, giving and scoring are more standard and objective, controls are better, and—experimentally if not clinically—the results may be considered more satisfactory.

The first problem of the present paper is, then, to determine whether a certain group scale, which was developed and standardized

¹Studies from the Psychological Laboratory of Indiana University.

²Research Associate in Psychology in Indiana University.

³See Sidney L. Pressey and Luella W. Cole, Irregularity on a Psychological Examination as a Measure of Mental Deterioration, Jr. of Abnormal Psychology, December, 1918; and Are the Present Psychological Scales Reliable for the Examination of Adults?, An Analytical Comparison of Examinations from Children and from Adults, Jr. of Abnormal Psychology, February, 1919; also Irregularity on a Binet Examination as a Measure of Reliability, Psychological Clinic, June, 1919.

in the public schools⁴ might be considered to give a satisfactory measure of the intelligence of adults and of atypical individuals of delinquent tendencies. In this connection an effort is made to find out whether a definite measure may not be obtained which will give a numerical expression of the extent to which such a scale is applicable to a given group; this problem is of the greatest importance, particularly in working with adults, and with children from homes where English is not spoken, or in environments which are otherwise unusual.

If, however, a scale is found not applicable to a certain group of cases the question at once arises as to why this may be so. The second, analytical, problem of the paper is, then, to determine in what ways and on what tests the results from these atypical groups deviate from the standard findings. These two problems will be taken up in order.

II. MATERIALS OF THE PRESENT STUDY

The materials of the present study were obtained with a brief group scale for measuring general intelligence recently developed at Indiana University. The scale has been described in detail in previous papers⁵ and need have only brief description here. It may be said, shortly, that the first test consists of twenty-five disarranged sentences, such as

“John broken window trees has the.”

Each sentence contains one superfluous word; this word the children are to find and cross out. The test may be considered to measure language ability and ingenuity of the verbal type. The second test consists of twenty-five lists, each list containing five words, such as

coat, shoes, hat, gloves, sail.

The children are to find the one thing in each line which does not belong with the other four, and cross it out. The test may be con-

⁴As are practically all the scales for measuring general intelligence now in use for work with adults and psychopathic cases. It often seems to be forgotten that both the Stanford-Binet and Yerkes Point Scale were standardized primarily on grade school children, and that neither one of these two scales have standards derived from unselected groups for ages above 12. The various group scales now in use, with the exception of the army scales, are also “school” scales. And as the writer hopes to show, the army scale Alpha has altogether too much of the scholastic element in it.

⁵Pressey, S. L., A Brief Group Scale for Use in School Surveys, Jr. of Educational Psychology, February, 1920; also, Cross-out Tests, with Suggestions as to a Group Scale of the Emotions, Jr. of Applied Psychology, June, 1919; and, School Surveys by Means of Group Tests of Intelligence, Sixth Annual Conference, on Educational Measurements, Indiana University, 1919.

sidered to involve information and practical judgment. The third test is made up of twenty-five lines of figures, such as

2 4 6 7 8.

The children are told that in each line the numbers are arranged according to some rule (in the example given, the rule, of course, is to count up by twos), but that in each line there is one number that breaks the rule. The children are to find this number, and cross it out. The test has been called a measure of arithmetical ingenuity. The last test consists of twenty-five lists, each containing five words, such as

dullness, foolishness, laziness, weakness, poverty.

The children are told simply to "cross out in each line the thing that is worst." The test may be considered to involve both vocabulary and moral discrimination.

It is evident that the four tests of the scale sample rather widely different phases of ability. And—a fact more important for the present study—the tests should be influenced in widely varying degrees by special circumstances of environment, and maturity. It might be expected that with maturity greater knowledge of the facts included in tests 2 and 4 would be acquired. Test 3 would be most closely associated with immediate schooling. Test 1 should give distinctive results as to familiarity with and ease in handling of the written language, and so should indicate the general level of literacy—and also prove especially interesting, perhaps, in work with groups including the foreign born. It should also be mentioned that these tests may be considered distinctly representative of the general run of group tests of intelligence now being used. In fact, this cross-out scale is the result of long and careful analytical study aiming at the selection of the best in such scales for purposes of rapid survey and research work. Thus the army scale Alpha contains two tests of arithmetic, one the same in general problem as test III. Test I of the cross-out scale is an improvement of the army disarranged sentence test. The second test is, in its emphasis upon practical information, largely similar to two tests of the army scale. And the moral discrimination test is the result of an effort to combine, in group test form, the merits of the Terman-Childs vocabulary test and the Binet "definition of moral terms" test.⁶ The cross-out series has the special merit, however, that the test form is much less artificial and complicated than is usual in

⁶See Terman, L. M., *The Measurement of Intelligence*, Houghton Mifflin Co., 1916.

the present-day group scale. The special groups with which the present paper deals should thus find it easier to express themselves on the cross-out tests than on most other forms—and if the cross-out scale is not applicable, it may be inferred (the writer believes) that the majority of such scales are less so.⁷

The norms for the cross-out test may be considered reasonably adequate and representative. The total number of cases involved is about 5,500. The cases are largely from Indiana, but the data includes results from Massachusetts and New York as well. The materials have been handled in such a way that the findings may be considered fairly representative of the total school population of these ages and grades. Reasonably satisfactory norms for each test have also been worked out. These various standards form the basis for the comparisons. The atypical groups are two in number. The first group consists of the entire population of the Indiana State School for Girls. The girls were from 11 through 19 in age; a total of 358 cases were tested. The second group consists of the attendants at a certain state hospital for the insane in Indiana; 57 cases were examined. The group may be considered not unrepresentative of the average small town adult population of southern Indiana.⁸

The question now is—first of all, of course, as to the showing made by these special groups on the tests—as to their general intelligence, one is almost tempted to say. But the fundamental questions are, as the writer sees them: (1) To what extent are such tests and scales applicable to such groups and (2) in what ways do such special groups give results differing from the standard findings?

III. RESULTS. (a) DIFFERENCES IN "GENERAL INTELLIGENCE"

With regard to the general composition of the population of the Girls' School it may be said that the girls are sent there by the courts as incorrigible, or for related reasons suggesting the need for special training under supervision. At the school they receive two types of training. The work in the "school of letters" is, in general, similar to the ordinary public school curriculum. Besides this there is the industrial training, almost wholly along the line of domestic arts. Most of the girls enter the school between 14 and 16, and require about two

⁷For the data regarding the derivation of the Cross-out Test see the previous papers already referred to, and also "The Practical Efficiency of a Group Scale of Intelligence," *Jr. of Applied Psychology*, March, 1919.

⁸The writer wishes to express his obligations to Dr. Kenosha Sessions, Head of the Indiana Girls' School, for her kindness and co-operation in the work, and to Miss Hazel Hansford, Psychologist, Southeastern Indiana Hospital for the Insane, for the data from the attendants.

or three years to complete their training—their progress being dependent partly on the quality of their work, but partly also on their conduct. There is a definite provision that feeble-minded girls shall not be admitted. Cases must be discharged when they reach twenty. The median age of the entire school is 17, and 61 per cent of the cases are either 16, 17 or 18. The median school grade is the sixth, and 72 per cent were in either the 5th, 6th or 7th grade—or had been when they left school. In general, the girls are kept in the "school of letters" until they have gone as far as they seem capable of going, or have finished the grade work.⁹

The standard median scores for each age and the median scores for the Girls' School run as follows:

Ages	10	11	12	13	14	15	16	17
Norm—								
No. Cases	1035	1034	1038	960	297	226	162	124
Median	39.8	46.3	52.1	58.2	94.6	69.9	76.6	80.7
Girls' School—								
No. Cases				12	23	47	72	79+56+59
Median				41.0	42.0	50.0	54.0	53.0

The above figures might suggest that the Girls' School cases were, on the average, about four years retarded mentally. It must be mentioned at once, however, that the norms above thirteen are *not* true norms for the entire population of these ages. They are norms only for the school population. And it is increasingly true for the norms for ages 15, 16 and 17 that the school group is a *selected* group, is the

⁹The scores of these cases when grouped according to their grade compare very favorably with the results obtained from the public school children. The norms by grade for the cross-out scale are compared with the results from the Girls' School in the following table; the results are in terms of median score in each case:

Grade	4	5	6	7	8
Median (public school children).....	35.7	43.7	51.4	57.4	65.2
Median (Girls' School).....	29.3	42.5	51.4	56.0	63.0

These results would perhaps suggest that the girls at the reform school were quite as able as the average school child. However, in dealing with grade norms for any type of test it is exceedingly important that information be available with regard to the age-grade distribution of the schools from which these norms were obtained. The median ages for the standard group above, in comparison with the median ages for the girls' school, are as follows:

Grade	4	5	6	7	8
Median (public School children).....	10.12	11.16	12.17	13.00	13.92
Median (Girls' School).....	16.25	16.57	17.11	17.75	18.01

These high median ages are, of course, due partly to the presence of some girls who have finished school; but the important element is the much greater retardation among these girls—a retardation which is fundamentally an effort to compensate for their poor average ability. For further discussion of such compensations as they appear in public schools see Pressey, S. L., *A Comparison of Two Cities and Their School Systems by Means of a Group Scale of Intelligence*, Educational Administration and Supervision, February, 1919.

select group of those who go to high school. Because the 17-year-old median for the Girls' School is 27 points below the standard median for that age is not, therefore, a valid reason for inferring that the girls at the School are markedly below average in ability. The norm for thirteen years is, however, based on an unselected group; practically all thirteen-year-old children are in school, and so are tested in such surveys as those on which the norms are based. We may believe that the median score for all thirteen-year-old children is close to 55. There is also some evidence to show that the mental age for the average adult population of the country is 13 years—that the “upper life age limit” in the growth of intelligence of average individuals is 13.¹⁰ It will be seen that the Girls' School medians are, however, distinctly below the 13-year median. The results can, perhaps, be more satisfactorily expressed in terms of the per cent of the Girls' School population scoring above the median for 13. Thirty-eight per cent of the girls thus score, and in contrast to this finding it is interesting to compare the older public school children. If we consider only the cases 15, 16 or 17 years old, in the Girls' School and in public school, it appears that eighty-eight per cent of the public school children score above the 13-year median, and that thirty-one per cent of the Girls' School cases, of these three ages, thus score.

Before any conclusions are drawn from these figures, however, the scores made by the hospital attendants should be considered. These attendants average around thirty years in age; they range in education from a minimum of country school to high school training. They come mostly from an Indiana town of a population of about 6,000 and the country near by. Their median score (55.6) is, it is very interesting to note, almost exactly the score for 13 years.¹¹ And, if we may take these results at their face value, we have definite evidence of the selected nature of high school students and of an inferiority, relatively slight, however, among the Girls' School cases.

IV. RESULTS. (b) DIFFERENCES IN “APPLICABILITY”

However, *can* these results be taken at their face value? And is it not possible to find some measure which will indicate definitely the extent to which the results obtained from these different groups may be considered sufficiently congruous with the results obtained from

¹⁰The conclusions are based largely on the psychological work in the army. See Doll, E. A., *The Growth of Intelligence*, Jr. of Educational Psychology, December, 1919.

¹¹It should be explained that in all the age tabulations, age at last birthday has been used; the norms are thus really for 12.5, 13.5, etc.

the standard group, in make-up of examination, for those standards legitimately to apply?

The writer has been for some years much interested in the problem as to whether "irregularity" on a Binet scale may not have some such significance. Careful study of the matter seemed to indicate that a statement of "reliability" might be thus obtained; illiterates, insane cases, very aged individuals, for the most part scattered widely on the Binet Scale. And it seemed quite evident that to these cases the scale was not, to a very high degree of nicety at least, applicable. Cannot, now, an analogous measure be derived from group scale results?

A very simple statement of this sort may be obtained by simply finding the difference between the highest and lowest score made by each individual on the separate tests of the scale. The norms for the four tests progressed with a high degree of similarity from age to age; thus the medians for 13 years for the four tests, in order, are as follows: 14.7, 15.2, 14.0, 15.1. In so far as a given thirteen-year-old child is a typical thirteen-year-old child, one would expect him to show a relatively slight deviation from these medians in his scores on the four tests. If, however, he is unusual in the make-up of his abilities, if (to take an actual case) a girl had never been to school, and so had never had any drill in formal arithmetic, but is an omniverous reader, one may expect a very low score on the third test of the scale (arithmetical ingenuity) and a very high score on the first and last test (verbal ingenuity and moral judgment and vocabulary). The result would be a large difference in score between the lowest and highest scores—in this instance, 18 points. As a matter of fact school children show a median distance between lowest and highest scores of 6.7 points. The Girls' School cases give a similar irregularity of 8.0 points and the irregularity of the hospital attendants is 10.5 points. There is here a suggestion that the composition given by the two special groups is different from the composition of the standard examinations. And an analysis is at once suggested in order to find out the nature of this difference.

In this analysis by test the comparison has again been on the basis of the 13-year norm and public school children 15, 16 and 17 years old have been contrasted with Girls' School cases of the same age and hospital attendants. And it may be said, shortly, that the "per

cent in each group scoring above the median for 13 years for each test" run as follows:

Tests	1	2	3	4
Public School	86	89	82	87
Girls' School	36	44	15	58
Hospital Attendants	47	53	19	69

That is, the score turns out to be a merging of a diversity of indications. And a mental age statement, at least in the case of the hospital attendants, would seem altogether unwarranted. The high school boys and girls have very little formal arithmetic. But for some reason, perhaps because they are used to academic and artificial problems, they deal with the arithmetical test very well. The two other groups fall off astoundingly on this same test. The special groups do best in the two tests involving information. Curiously enough, these two last groups do relatively best on the test of moral discrimination!

V. THE INAPPLICABILITY OF SCALES DESIGNED FOR USE WITH SCHOOL CHILDREN IN DEALING WITH ADULTS

Well, this variability from test to test in performance is surely surprising enough. It should be pointed out at once that this can hardly be considered a defect of this particular scale. As was said in the beginning, each one of the four tests has a long and honorable history; the cross-out form simply permits more ready giving and scoring. The test of arithmetical genuity is taken directly from the army scale Alpha, but with a shift in form, which makes the test easier to take, since writing is eliminated. The fourth test is primarily a measure of vocabulary (the moral choices are for the most part fairly obvious, and mistakes often occur simply because the children are not familiar with the words). Professor Terman tells us that the vocabulary test is probably the most valuable single test of the Binet scale. Yet the relative standing of these three groups on these two tests differs surprisingly, and it is surely putting a strain upon one's credulity to suppose that one can, in combining such scores, by some mysterious alchemy produce from such divergent elements a fundamental measure of a unitary general intelligence.

These results are, of course, based on a relatively small number of cases. But they are quite like some other results which the writer has found; he sees no reason to question the general trend of the findings. The fundamental fact is, he believes, that in a great deal of the present work, investigators have been altogether too much dominated

by the concept of general intelligence, have missed the richness of their data as a consequence, and have found what they went after—a generalized total of a very vague significance, but appealing largely to the imagination. It is high time that careful analytical studies of results given with various types of tests, with various types of cases, were made. And it is with the hope that further impetus may be given to such work that the present results are being published.

SUMMARY

The paper may be briefly summarized. It deals with a comparison of public school children, girls at a state reform school, and attendants at a state hospital for insane. Comparisons are for the most part with median score of public school children 13 years of age. It is found that

(1) Public school children 15, 16 and 17 years old show 88 per cent scoring above the 13-year norm, on a group scale of intelligence. Thirty-one per cent of the Girls' School cases thus score, forty-five per cent of the attendants.

(2) As a measure of "irregularity" distance between the highest and lowest scores on the individual test is taken. Excessive irregularity is found for the Girls' School and a very marked irregularity for the hospital attendants. It is concluded that the composition of the examination yielded by these last two groups must be different from the composition of the examinations yielded by the standard group of public school children.

(3) Analysis by test shows striking differences between the groups, the Girls' School and hospital groups differing chiefly in an extremely poor performance on the arithmetical test and a relatively good performance on a test of vocabulary.

(4) It is argued that, in considering results obtained by scales measuring "general intelligence" from such special groups, conclusions regarding the general abilities of these groups should not be drawn until analysis by test has shown the composition of the examinations to be sufficiently analogous to the make-up of the standard examination to warrant such inferences.